American Fern Journal 100(1):1–15 (2010)

Cloning and Sequence Analysis of Red/Blue Light Chimeric Photoreceptor Genes from Three Fern Species (*Coniogramme intermedia* var. *glabra*, *Plagiogyria distinctissima* and *Pronephrium lakhimpurnense*)

YONGXIA YANG

Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, China; Graduate School, Chinese Academy of Sciences, Beijing 100039, China; Henan Agricultural University, Zhengzhou, Henan 450002, China

XIAOQIONG QI and LIN SEN

Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, China

YINGJUAN SU*

State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, Guangdong 510275, China

TING WANG*

Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, China

ABSTRACT.—Neochrome (NEO) is a particular chimeric photoreceptor that can respond to red/far red as well as blue light. Using this photoreceptor, some cryptogams may enhance light sensitivity at low light environment. However, the *NEO* sequence information is extremely limited. Using inverse PCR and single oligonucleotide nested PCR genome-walking approaches, the full-length genomic sequences of *NEO* genes were determined for three fern species *Coniogramme intermedia* var. *glabra*, *Plagiogyria distinctissima* and *Pronephrium lakhimpurnense*. Sequence and structure analysis results indicate that they possess a conserved structure and organization. Introns are absent in three genes. They each contain an 4332, 4308 and 4317-bp ORF (Open reading frame), encoding a deduced protein with 1443, 1435 and 1438 amino acid residues. Their N-terminus consists of a putative functional phytochrome sensory module including PAS (Per / Arnt / Sim), GAF (cGMP specific phosphodiesterase / adenylate cyclases / formate hydrogen lyase transcription) and PHY (phytochrome) domains, whereas the C-terminus contains a nearly complete phototropin with two LOV (Light, Oxygen and Voltage-sensing) and one STKc (Serine/Threonine protein kinases, catalytic) domains. Phylogenetic analyses revealed that they are evolutionarily close to their homolog *AcNEO* found in *Adiantum capillus-veneris*. The 3D structures of the functional domains were computer-modeled, and the pivotal amino acid sites for each functional domain were also identified. This study represents the first comparative analysis of *NEO* nucleotide sequences and structures in ferns.

KEY WORDS.—Fern, Red/blue light chimeric photoreceptors, Sequence analysis, Structure prediction

Light is one of the most essential environmental factors for plants that perform photosynthesis. Plants have developed a variety of photoreceptors

* Corresponding author: Yingjuan Su, Tel: +86-20-84035090, Fax: +86-20-84036215, e-mail: suyj@mail.sysu.edu.cn; Ting Wang, Tel: +86-27-87510677, Fax: +86-27-87510251, e-mail: tingwang@wbgcas.cn.

that perceive light signals from UV-B to far-red regions (Pepper *et al.*, 2001; Quail, 2002). Phytochromes predominantly absorb red and far-red light, using a bilin (or linear tetrapyrrole) chromophore for light detection in seed plants and in cryptogams (Quail, 2002), and phototropins are UV-A/blue light receptors (Briggs and Christie, 2002; Folta and Spalding, 2001). Many photomorphogenic responses mediated by these photoreceptors, for example, circadian clocks, shade-avoidance, phototropism, stomatal opening and chloroplast relocation, have been described (Christie, 2007; Rockwell *et al.*, 2006). Most plant species from algae to flowering plants use blue light for inducing phototropism and chloroplast movement (Campbell and Liscum, 2001; Sakai *et al.*, 2001), but some ferns and green algae use red as well as blue light for the regulation of these responses (Kagawa and Wada, 1994; Wada and Kadota, 1989). In the case of *Adiantum capillus-veneris* L., a particular chimeric photoreceptor named phytochrome 3 (later renamed as neochrome, NEO) has been identified (Nozue *et al.*, 1998; Suetsugu *et al.*, 2005), which can sense red and blue light (RL and BL) simultaneously to regulate phototropism and chloroplast movement (Kanegae *et al.*, 2006; Kawai *et al.*, 2003).

Functional domains of NEO have been characterized through biochemical, genetic, and bioinformatic approaches (Crosson and Moffat, 2001; Pandini and Bonati, 2005; Tokutomi *et al.*, 2008; Wagner *et al.*, 2005). Its N-terminus is quite similar to canonical phytochromes, whereas the C-terminus is made up of a complete phototropin motif. It has been suggested that both phytochrome and phototropin photosensory systems coexist in NEO, conferring it the potential to process red and blue light signals synergistically, which increases the light sensitivity (Kanegae *et al.*, 2006). Many polypod ferns preferentially grow in shady habitats under the canopy of dense angiosperm forests. For them, this dual-channel photoreceptor has been proposed to play a central role in developing adaptive mechanisms and promoting species diversification under low-light angiosperm forest niches (Kawai *et al.*, 2003; Schneider *et al.*, 2004).

To date, similar NEO-like chimeric photoreceptors have only been identified in a few cryptogams; two were isolated from the green alga *Mougeotia scalaris* Hass. (Suetsugu *et al.*, 2005), and four from ferns (one full-length sequence from *Adiantum* and three short fragments from *Dryopteris filix-mas* (L.) Schott, *Onoclea sensibilis* L. and *Hypolepis punctata* Mett., respectively) (Kawai *et al.*, 2003). Thus, *NEO* sequence information is still extremely limited. Here we report the isolation and sequence analyses of the full-length genomic *NEO* sequences from two polypod ferns *Coniogramme intermedia Hieron.* var. *glabra* Ching, *Pronephrium lakhimpurnense* (Ros.) Holtt., and one tree fern *Plagiogyria distinctissima* Ching. Phylogenetic analyses were conducted to discern the gene origin, and the secondary and tertiary structures of deduced proteins were further predicted using bioinformatics methods.

Materials and Methods

Plant materials and DNA isolation.—All ferns were cultivated at Wuhan Botanical Garden, Chinese Academy of Sciences, Hubei, China. For each plant,

TABLE 1. Primers used in diversified genome-walking experiments.

Methods	Oligo name[a]	Sequence (5′→3′)
IPCR	5′ GDP F1	ATCACGCCCAATTCCGAY(A/C/T)TTC
	5′ GSP R1	CACAGCATTGTCTGCGTCATAAG
	5′ GSP F2	GTTCTGGTTCAGGTCGCATAC
	5′ GSP R2	AGAAGACCCCACCAGGCTAATG
Universal PCR	5′ GSP F3	TCAGCAATGTCCGCCAAGACGAAG
	5′ GSP R3	TAGGCCATCACCCGATCGTAGC
	GSP F1	GATGGCCTACAAGTTCCACGAG
	GSP R1	TGGCAGGGGTAGTGATAGA
	GDP F2	ATCACGCCCAATTCCGAY(A/C/T)TTC
	GDP R2	TGH(C/T)ACTCCGATGAAGTACTGGA
	GSP F3	GAGGTGCTCGGTGAGAACTG
	GSP R3	CTCGGGCTTGAGGTCTCGGTAG
SON	3′ GSP 1	TGAGGAATCTGCCCGCTTCTACGC
	3′ GSP 2	CCACTGCATGGGTGTCATCTAC
IPCR	3′ GSP F2	TCAAGCCCGAGAACCTGCTACTG
	3′ GSP R2	ATCCCTCCGTGTACTTGCTAACC
	3′ GSP F3	GAGGACTGGTGGGCGTTGGGAATC
	3′ GSP R3	CGGTAAGATTGAAGAACCCTGTG

[a] GSP referred to genomic specific primer, while GDP referred to degenerated primer.

three pieces of fresh leaves were collected to isolate genomic DNA following modified CTAB protocols (Sambrook *et al.*, 2001).

Cloning and sequencing of full-length NEO gene from C. intermedia var. *glabra.*—Full-length genomic *NEO* genes were amplified using diversified genome-walking PCR approaches. Primers used are listed in Table 1. Specific and degenerate primers were designed to amplify the internal *NEO* regions of *C. intermedia* var. *glabra* based on the conserved phytochrome chromophore-binding and phototropin regions of between *Adiantum NEO* (Accession No.: AB012082) and *M. scalaris* homologous sequences (Accession No.: AB206967, AB206966). Using the primer pairs GSP F1/R1, GDP F2/R2 and GSP F3/R3, universal PCR amplifications were performed in 20 μl of reaction mixture containing 50 mM KCl, 10 mM Tris–HCl (pH 9.0), 0.1% Triton X-100, 2.0 mM $MgCl_2$, 0.2 mM each deoxynucleoside triphosphate, 1 U *Taq* DNA polymerase, 0.4 μM primers, 50 ng genomic DNA, and DNA-free water. PCR amplifications were carried out under the following conditions: the template was denatured at 94°C for 5 min followed by 35 cycles of amplification (94°C for 30 s, 62°C for 30 s and 72°C for 90 s) and a final elongation step at 72°C for 10 min.

Meanwhile, Inverse PCR (IPCR; Huang, 1997) was used to obtain the 5′-portion of the gene. The genomic DNA of *C. intermedia* var. *glabra.* was digested with ApoI restriction enzymes and ligated with T4 DNA ligase. Then the ligation solution was used as the PCR template. The first round and nested PCR amplification were performed with primer pairs 5′ GDP F1/5′ GSP R1, 5′ GSP F2/5′ GSP R2 and with LA *Taq*E following the procedures the same as described above except that the annealing temperature was 56°C.

The 3′ *NEO* sequence of *C. intermedia* var. *glabra* was cloned by single oligonucleotide nested PCR (SON PCR; Antal *et al.*, 2004) using primers 3′ GSP 1 and 3′ GSP 2 (Table 1) for the first and second round PCR, separately. The DNA template for the first round PCR was 50 ng of genomic DNA. The PCR program was: one denaturation step for 3 min at 94°C, five cycles of amplification (30 s at 94°C, 1 min at 60°C, 3 min at 72°C) followed by one ramping step (30 s at 94°C, 3 min at 29°C, 3 min ramp to 72°C, 2.5 min at 72°C) and 60 new amplification cycles (30 s at 94°C, 1 min at 60°C, 2.5 min at 72°C) with a 10-min extension step at 72°C. The second round PCR used 2 μl of a 1:50 dilution of the first round PCR product as the template, and was performed with the following conditions: 5-min denaturation at 94°C, 30 cycles of amplification (30 s at 94°C, 1 min at 60°C, 2.5 min at 72°C) and 10-min terminal extension at 72°C.

All PCR products amplified by the above steps were recovered from 1% agarose/TAE gels and purified with a quick PCR Purification Kit (Promega), and then cloned into PMD19-T vectors (Takara). The plasmids were transformed to *Escherichia coli* strain DH5α. Based on the color reaction using Xgal-IPTG system and PCR identification, white transformants were selected and positive clones were sequenced with an ABI PRISM 3730 DNA analyzer. Three clones were sequenced for each amplicon to control *Taq* polymerase errors. The overlapping sequences from various amplification steps were assembled in a single contig.

Isolation of NEO genes from P. distinctissima and P. lakhimpurnense.—The internal regions of *NEO* genes for *P. distinctissima* and *P. lakhimpurnense* were amplified by using the same primers and PCR procedures as *C. intermedia* var. *glabra*. Primers 5′ GSP F3 and 5′ GSP R3 were used to obtain the 5′ regions. The former was designed based on the 5′ conservative *NEO* sequences between *A. capillus-veneris* and *C. intermedia* var. *glabra*, while the latter according to the internal regions obtained above. The 3′ *NEO* region of *P. distinctissima* was obtained using the same SON PCR procedures and primers as those of *C. intermedia* var. *glabra*, whereas the 3′ sequence of *P. lakhimpurnense NEO* was gained by IPCR with ApoI restriction enzyme system and the primer pairs 3′ GSP F2/ R2 and 3′GSP F3/ R3. Amplified PCR products were purified, cloned, sequenced and assembled as described above.

Sequence accession numbers.—*NEO* gene sequences were submitted to NCBI database with the Accession Numbers FJ501964–FJ501966.

Sequence analysis and prediction of the NEO *structure.*—The open reading frame (ORF) finder and BLAST tool were used to predict coding sequence and estimate similarities of *NEO* with other genes. The translated amino acid sequences of each gene were analyzed using the program Protparam on Expasy website. Protein domain search, prediction of protein hydrophobicity, transmembrane structures and signal peptide cleavage sites were performed with Prosite (Hofmann *et al.*, 1999), ProtScale (Gasteiger *et al.*, 2005), TMpred (Hofmann and Stoffel, 1993) and SignalP (Emanuelsson *et al.*, 2007) programs, respectively. SOPMA (Geourjon and Deleage, 1995) analyses were carried out to predict the secondary structure. Three-dimensional (3D) structures of

functional domains were computer-modeled by SWISS-MODEL (http://www.isb-sib.ch/) program. 3D structures of PAS (Per / Arnt / Sim) and GAF (cGMP specific phosphodiesterase / adenylate cyclases / formate hydrogen lyase transcription) were predicted based on the 1.45 Å crystal structure of *Deinococcus radiodurans* Phytochrome (PDB code: 2o9c). The LOV (Light, Oxygen and Voltage-sensing) 1, LOV2 and STKc (Serine/Threonine protein kinases, catalytic) domains were computer-modeled using fungal photoreceptor VVD (PDB code: 3d72A; Resolution: 1.65 Å), *Adiantum* NEO LOV2 (PDB code: 1jnuB; 2.60 Å) and Human Myotonic Dystrophy Protein Kinase (PDB code: 2vd5A; 2.80 Å) as templates, respectively. Models were displayed with the PyMol program (DeLano, 2002).

Sequence alignment and phylogenetic analysis.—Other than the sequences cloned in this work, sequences used for comparison and phylogenetic analysis were downloaded from Genbank. Complete species names and accession numbers are listed in Table 2. Based on the predicted structure of NEO, we divided the *NEO* sequences data into N-terminal phytochrome-like and C-terminal phototropin-like regions. Sequences were aligned using Clustal X 2.0 and then adjusted manually. The phylogenetic trees were constructed from the aligned coding DNA using the maximum likelihood (ML) method as implemented in PHYML (Guindon and Gascuel, 2003). Before ML analyses, we first identified the best fit model of molecular evolution for the N- and C-terminal datasets using the Akaike Information Criterion (AIC) as implemented in ModelTest (Posada, 2003). The model of the general time reversible (GTR) + 4G (four categories of Gamma substitution rates) + I (invariable sites) model were applied to each datasets. A nonparametric bootstrap test was conducted with 1000 replicates.

Results and Discussion

Cloning and sequence analysis of the complete coding sequences.—Using diversified genome-walking PCR approaches and the primers in Table 1, the *NEO* fragments were amplified from three fern species (Fig. 1). After cloning and sequencing, the assembling results displayed that the *NEO* genes from three fern species, *C. intermedia* var. *glabra*, *P. distinctissima*, and *P. lakhimpurnense*, are 4937 bp, 4424 bp, and 4456 bp in length, respectively. Comparative analyses with the full-length *A. capillus-veneris* L. *NEO* (*AcNEO*) cDNA sequence showed that introns are all absent in these genes. Ahead of the original codon ATG and after the stop codon TGA, there are 294-bp, 6-bp and 6-bp 5′- untranslated regions (UTR), in addition to a 311-bp, 100-bp and 139-bp 3′-UTR for *C. intermedia* var. *glabra*, *P. distinctissima*, and *P. lakhimpurnense*, respectively. Each sequence consists of an ORF of 4332 (*C. intermedia* var. *glabra*), 4308 (*P. distinctissima*) and 4317 bp (*P. lakhimpurnense*), which is very similar to that of *AcNEO* (4398 bp). Their guanine–cytosine (GC) contents are 59.5%, 61.0% and 58.7%, respectively. The three genes each encode a deduced protein of 1443, 1435 and 1438 amino acid residues with a

TABLE 2. Species names and accession numbers of nucleotide acid sequences used for comparison and phylogenetic analysis.

Species	Gene names	Accession numbers	Species	Gene names	Accession numbers
Adiantum capillus-veneris	AcPhy1	AB016168	*Pinus sylvestris*	PsPhyP	EU203168
Adiantum capillus-veneris	AcPhy2	AB016232	*Sorghum bicolor*	SbPhyB	AY466087
Arabidopsis thaliana	AtPhyA	X17341	*Selaginella martensii*	SmPhy1	X61458
Arabidopsis thaliana	AtPhyB	X17342	*Solanum tuberosum*	StPhyB	Y14572
Arabidopsis thaliana	AtPhyC	X17343	*Zea mays*	ZmPhyB1	AY234827
Arabidopsis thaliana	AtPhyD	X76609	*Zea mays*	ZmPhyB2	AY234828
Arabidopsis thaliana	AtPhyE	X76610	*Adiantum capillus-veneris*	AcPhot1	AB037188
Ceratodon purpureus	CpPhy1	U87632	*Adiantum capillus-veneris*	AcPhot2	AB115545
Ceratodon purpureus	CpPhy2	U72993	*Avena sativa*	AsNph1-1	AF033096
Ceratodon purpureus	CpPhy3	AY123149	*Avena sativa*	AsNph1-2	AF033097
Ceratodon purpureus	CpPhy4	EU122393	*Arabidopsis thaliana*	AtPhot1	NM_114447
Pomoea nil	InPhyE	U39787	*Arabidopsis thaliana*	AtPhot2	NM_180879
Lycopersicon esculentum	LePhyB1	AJ002281	*Chlamydomonas reinhardtii*	CrPhot	AJ416557
Lycopersicon esculentum	LePhyF	U32444	*Lycopersicon esculentum*	LePhot1	EF063359
Lotus japonicus	LjPhyB	AB264087	*Lycopersicon esculentum*	LePhot2	DQ886531
Mesotaenium caldariorum	McPhy	U31284	*Mougeotia scalaris*	MsPhotA	AB206963
Marchantia paleacea	MpPhy	AB022917	*Mougeotia scalaris*	MsPhotB	AB206964
Mougeotia scalaris	MsPhy1	AB206965	*Oryza sativa*	OsNph1	AJ252142
Mougeotia scalaris	MsNeo1	AB206966	*Oryza sativa*	OsNph1a	AB018444
Mougeotia scalaris	MsNeo2	AB206967	*Oryza sativa*	OsNph1b	AB018443
Nicotiana plumbaginifolia	NpPhyB	Y14676	*Physcomitrella patens*	PpPhotA1	XM_001774204
Neurospora crassa	NcPhy1	DQ128077	*Physcomitrella patens*	PpPhotA2	XM_001774562
Oryza sativa	OsPhyA	AB109891	*Physcomitrella patens*	PpPhotA3	XM_001765356
Oryza sativa	OsPhyB	AB183525	*Physcomitrella patens*	PpPhotA4	XM_001763052
Oryza sativa	OsPhyC	AF141942	*Physcomitrella patens*	PpPhotB1	XM_001766357
Picea abies	PaPhyO	U60264	*Physcomitrella patens*	PpPhotB2	XM_001785674
Populus balsamifera	PbPhyB1	AF309806	*Physcomitrella patens*	PpPhotB3	XM_001755269
Populus balsamifera	PbPhyB2	AF309807	*Pisum sativum*	PsPhot1	AY093427
Physcomitrella patens	PpPhy1	XM_001778103	*Phaseolus vulgaris*	PvPhot1a	AB204872
Physcomitrella patens	PpPhy2	XM_001782287	*Phaseolus vulgaris*	PvPhot1b	AB204873
Physcomitrella patens	PpPhy3	AB275306	*Phaseolus vulgaris*	PvPhot2	AB204874
Physcomitrella patens	PpPHY4	XM_001773498	*Vicia faba*	VfPhot1a	AB095909
Physcomitrella patens	PpPHY5a	XM_001761093	*Vicia faba*	VfPhot1b	AB095910
Physcomitrella patens	PpPHY5b3	XM_001767172	*Zea mays*	ZmPhot1	NM_001111416
Physcomitrella patens	PpPhy5c	XM_001754314			

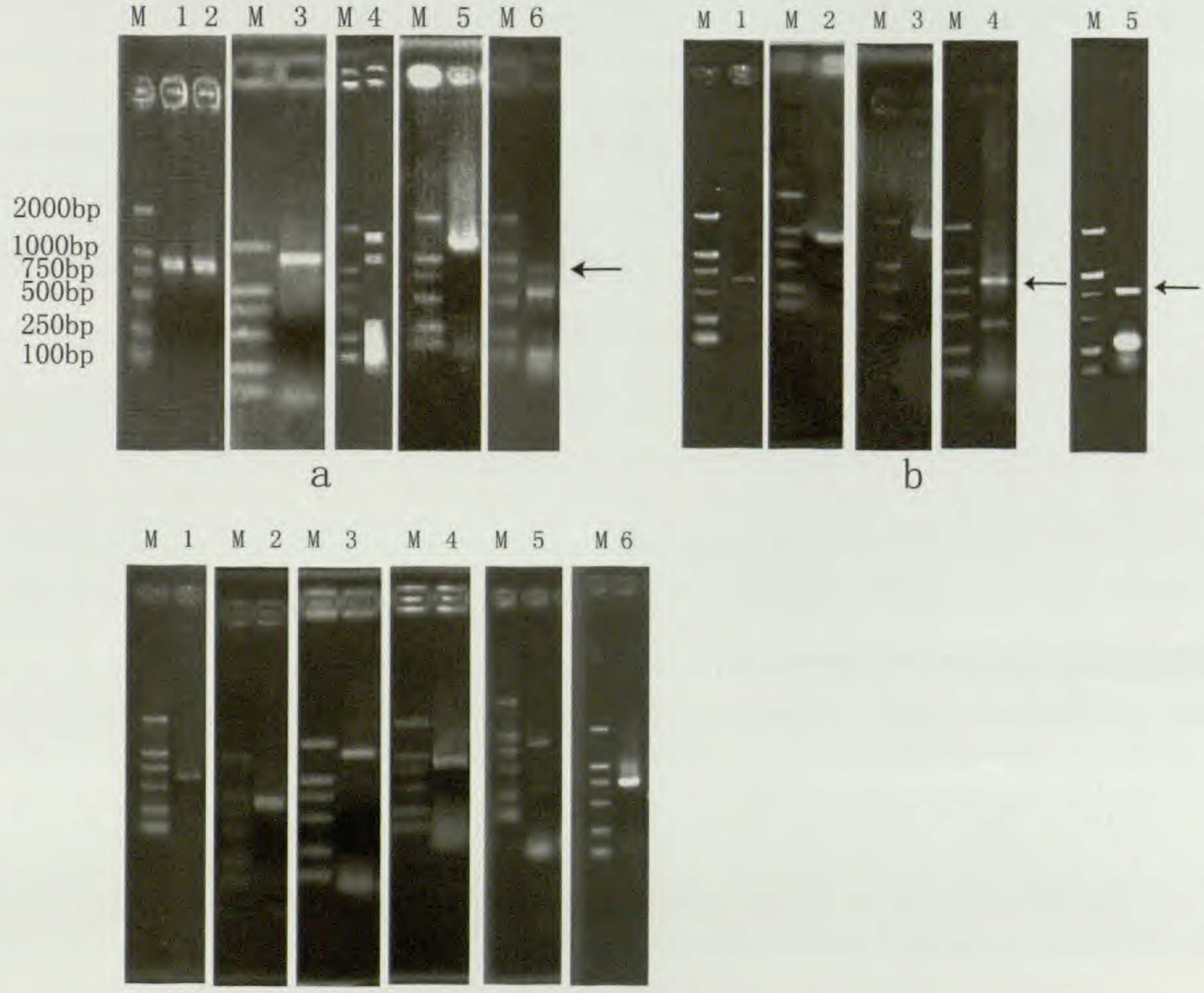

FIG. 1. Agarose gel electrophoresis of amplification products for *Coniogramme intermedia*. var. *glabra* **(a)**, *Plagiogyria distinctissima* **(b)**, and *Pronephrium lakhimpurnense NEO* fragments **(c)**. M is molecular weight marker (DL2000). **a1** and **a2:** the products amplified by the primer pairs GSP F1/R1and GSP F3/R3; **a3:** the products amplified by the primer pairs GDP F2/R2; **a4** and **a5:** the products amplified by the first and second IPCR using the primer pairs 5′ GDP F1/5′ GSP R1 and 5′ GSP F2/5′ GSP R2, respectively; a6: the amplification products of SON PCR by the primer 3′ GSP 2. **b1:** the products amplified by the primer pairs 5′ GSP F3/5′ GSP R3; **b2:** the products amplified by the primer pairs GSP F1/R1; **b3:** the products amplified by the primer pairs GDP F2/R2; **b4:** the products amplified by the primer pairs GSP F3/R3; **b5:** the amplification products of SON PCR by the primer 3′ GSP 2. **c1:** the products amplified by the primer pairs 5′ GSP F3/5′ GSP R3; **c2:** the products amplified by the primer pairs GSP F1/R1; **c3:** the products amplified by the primer pairs GDP F2/R2; **c4:** the products amplified by the primer pairs GSP F3/R3; **c5** and **c6:** the amplification products of the first and second IPCR by the primer pairs 3′ GSP F2/ R2 and 3′GSP F3/ R3, respectively.

theoretical isoelectric point (pI) of 6.44, 6.27 and 6.28 and a calculated molecular mass about 160.15, 158.51 and 158.28kDa.

Amino acid compositions of the three deduced proteins are very similar to each other. The richest amino acid is Leu (9.8~10.5% by frequency), followed by Ala (8.0~8.7%), Gly (7.8~8.7%), Ser (7.3~7.8%) and Arg (7.1~7.5%). Acidic and basic amino acids constitute about 12% and 11% of the protein separately. Of the total amino acids, 23% are polar while 32% charged. The

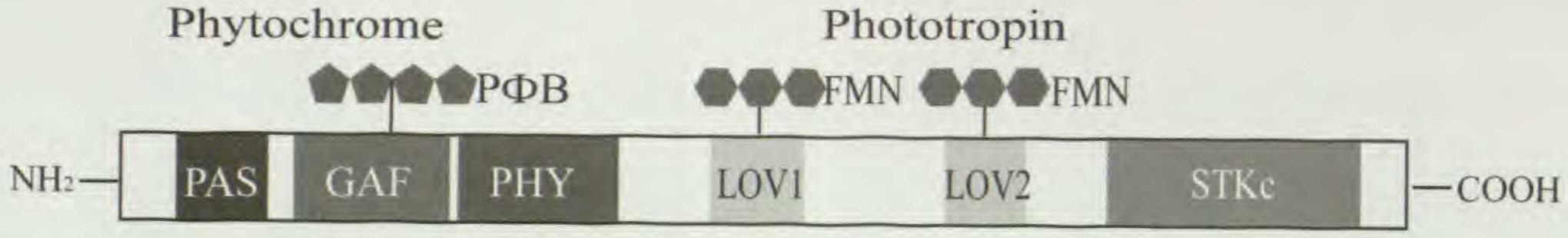

FIG. 2. The common structure of NEO in fern species. It shows that the N-terminal phytochrome chromophore domain contains PAS, GAF, and PHY domains and C-terminal phototropin motif possesses two LOV and one STKc domains.

deduced NEO proteins were predicted to have a net charge of −5.1, −9.2, and −7.9 at pH 7.0. They each contain 488, 495 and 494 hydrophobic amino acids (33.8%, 34.5% and 34.4% by frequency), and are all predicted to have a number of typical hydrophobic regions identified by Kyte and Doolittle hydropathy analyses (Gasteiger *et al.*, 2005). Each protein possesses several helices thought to be transmembrane domains with the N- terminal domains on the intracellular side. However, analysis of the primary structure did not reveal any signal peptide, indicating that they are most probably cytoplasmic. The results are in accordance with *Ac*NEO, which is a plasma-membrane localized protein (Nozue *et al.*, 1998). However, its putative association with the membrane remains unclear.

The three genes show 78~80% identity with *AcNEO* at the nucleotide acid level as well as >88% identity among themselves (Table 3). The N-terminal sections of *NEO* exhibits 40~58% identity with seeded vascular and non-vascular plant *PHY*s, and the C-terminal regions show 40~52% identity with plant *PHOT*s. Protein domain searches displayed that all three proteins possess the typical features of fern NEO family. They share three domains of phytochrome photosensory core region in the N-terminus and a nearly full phototropin homolog in the C-terminus (Fig. 2). Henceforth, we denoted the three new fern *PHY-PHOT* chimera genes as *C. intermedia* var. *glabra NEO* (*CiNEO*), *P. distinctissima NEO* (*PdNEO*), and *P. lakhimpurnense NEO* (*PlNEO*), respectively.

By contrast, the sequences of the chimeric photoreceptors are not conserved between the fern and algae. Fern *NEO*s are all distantly related to *Mougeotia scalaris NEO* (*MsNEO*)1 and *MsNEO*2 with <27% identity (Table 3). *MsNEO*1 and *MsNEO*2 exhibit lower identity to seeded vascular and non-vascular plant *PHY*s (<46%) and *PHOT*s (<45%) compared with fern *NEO*. In combination with the results that fern *NEO* genes are intronless but *MsNEO*1 and *MsNEO*2 have many introns (Nozue *et al.*, 1998; Suetsugu *et al.*, 2005), we also speculate that fern and alga *NEO* arose independently and evolved convergently. This is in agreement with a previous study (Suetsugu *et al.*, 2005). However, it cannot be excluded that the fern *NEO*s might have experienced intron losses during their evolution.

Phylogenetic analysis.—Fern and alga *NEO*s each constitute a clade with strong statistical support in the ML trees (Fig. 3). They further form a major clade in the tree established from the sequences of phytochrome sensor domains, while seeded vascular and non-vascular plants (including the

TABLE 3. Percentage of sequence identity among fern and alga *NEO*, canonical phytochromes, and phototropins.[a]

	AcNEO	*CiNEO*	*PdNEO*	*PlNEO*	*MsNEO1*	*MsNEO2*
AcNEO						
CiNEO	78.6					
PdNEO	79.5	89.4				
PlNEO	78.3	88.0	89.1			
MsNEO1	26.3	26.5	26.6	26.3		
MsNEO2	26.8	26.8	27.1	26.8	55.9	
Plant *PHY*[b]	41.2~56.8	41.1~57.2	41.2~57.6	40.7~56.1	40.6~45.5	39.9~45.5
Plant *PHOT*[b]	40.5~51.7	40.3~51.3	40.3~51.9	40.6~51.3	33.2~45.1	33.3~45.0

[a] The sequences of *CiNEO*, *PdNEO* and *PlNEO* were obtained by cloning and sequencing in this work, and the others were downloaded from Genbank database. Complete species names and accession numbers were listed in TABLE 2.

[b] The alignment with plant *PHY* and *PHOT* used the phytochrome-like N-terminal and phototropin-like C-terminal region, separately. Plant *PHY* and plant *PHOT* were in accordance with *PHY* and *PHOT* in molecular evolution analysis below.

angiosperm, gymnosperm, fern, moss, liverwort and alga) comprise another (Fig 3a). By contrast, for the ML tree constructed using the sequences of phototropin domains (Fig 3b), the fern *NEO*s were grouped together with seeded vascular and non-vascular plant *PHOT*s rather than alga *NEO*s. However, in each tree none of these clades received strong support. *NEO* has been regarded as the fusion between phytochrome and phototropin motifs (Suetsugu *et al.*, 2005). Since a robust phylogeny cannot be reconstructed here, we must await more necessary data to resolve the origin of fern *NEO* genes.

Predictions of secondary and tertiary structure.—The predicted secondary structures exhibit that the putative NEO polypeptides all contain about 37% alpha helices, 17% extended strands, 9% β turns and 37% random coils. Alpha helices and random coils constitute the interlaced domain of the main part of the secondary structure.

The predicted 3D structure of functional domains in *Ci*NEO, *Pd*NEO, and *Pl*NEO are displayed in Fig. 4. The 3D structure models of PAS and GAF domains were constructed with E-values of $2.9E^{-28}$, $1.8E^{-30}$, and $1.2E^{-29}$ for *Ci*NEO, *Pd*NEO, and *Pl*NEO, respectively. The PAS and GAF domains of the three proteins are highly conserved, which all share the characteristic structure of plant phytochromes (Wagner *et al.*, 2005). The main difference among *Ci*NEO, *Pd*NEO, and *Pl*NEO is that *Pd*NEO has a longer β-sheet at GAF domain, but *Ci*NEO and *Pl*NEO are random coils in that position. However, the principle of this difference remains unclear. Except for that, the three proteins each contain a figure-eight knot donated by GAF and PAS domains, and they all reserve the conserved cysteine residue (Cys306 for *Ci*NEO, Cys300 for *Pd*NEO and Cys301 for *Pl*NEO) in the GAF domain, which is associated with phytochromobilin (PΦB) by a covalent linkage (Boylan and Quail, 1991). The adjacent histidine residue (His307 for *Ci*NEO, His301 for *Pd*NEO and His302 for *Pl*NEO), required for tight chromophore-protein interaction, is also conserved.

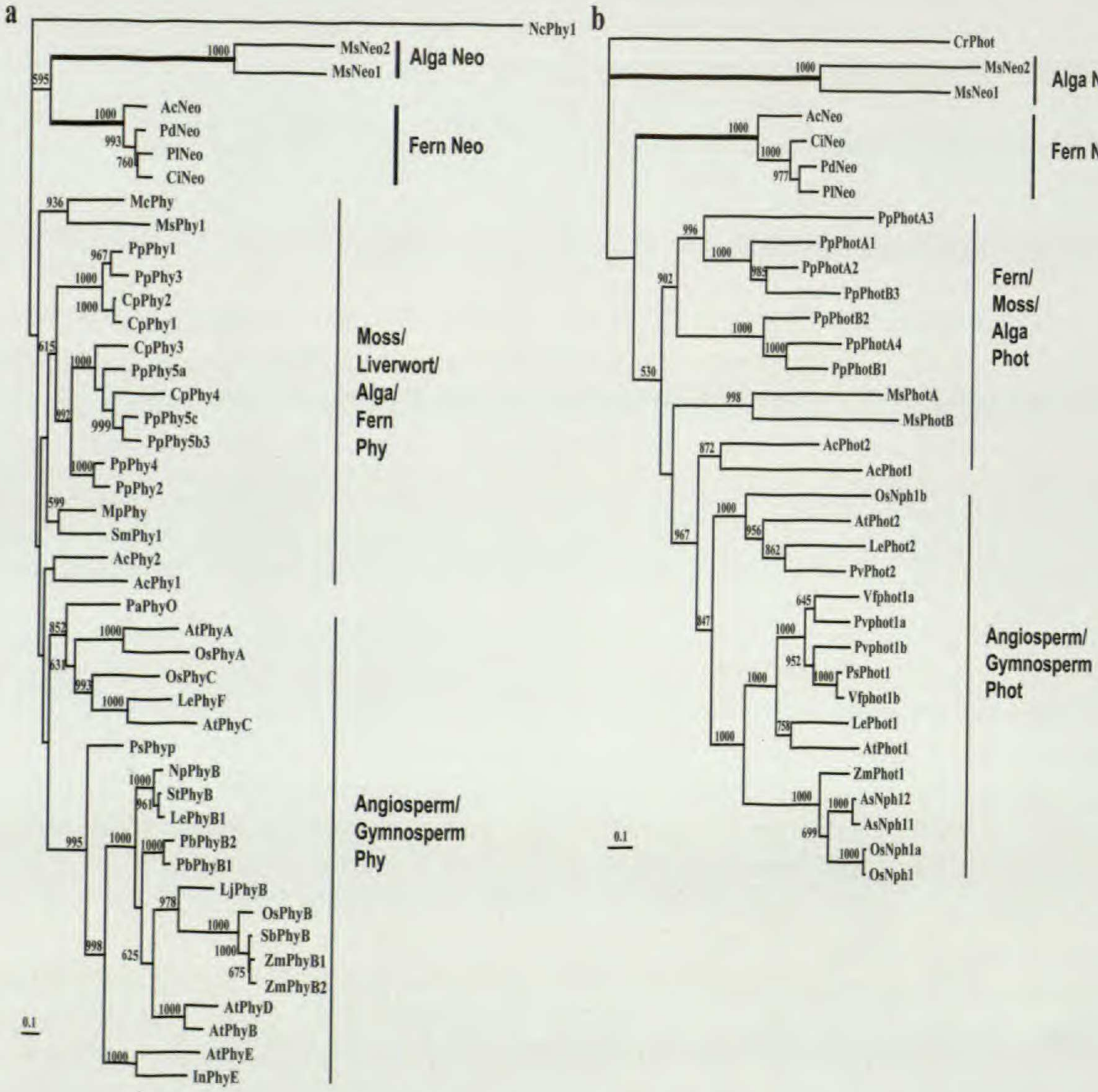

FIG. 3. Phylogenetic trees established from nucleotide acid sequences of phytochrome sensor domains **(a)** and phototropin domains **(b)** generated by the maximum likelihood method using PHYML software, which are arbitrarily rooted with *Neurospora crassa* phytochrome (NcPhy1) and *Chlamydomonas reinhardtii* blue light receptor (CrPhot). Numbers above branches or near to the right of nodes indicate bootstrap values (only values>500 are indicated) obtained from 1000 replications. The scale bar represents substitution number per nucleotide site. Neo branches are marked with bold dark lines.

The LOV1 and LOV2 structures of three NEO are nearly identical (Fig. 4b and 4c), which is consistent with the high degree of sequence conservation at the amino acid level (Fig. 5). The LOV domains all possess the typical α/β fold of the PAS family, comprising five stranded anti-parallel β-sheets ($\beta1$–$\beta5$) and three α-helical segments ($\alpha1$–$\alpha3$), which are with a backbone β-scaffold and a helical connector forming a tight cage enclosing the FMN (Pandini and Bonati, 2005). All LOV domains contain the highly conserved motif NCRFLQ and constitute a single turn of a 3_{10} helix (a′A), contributing to the chemistry of photoactivation and subsequent downstream signaling (Crosson and Moffat, 2001). The LOV domains also contain the conserved flavin-binding residues

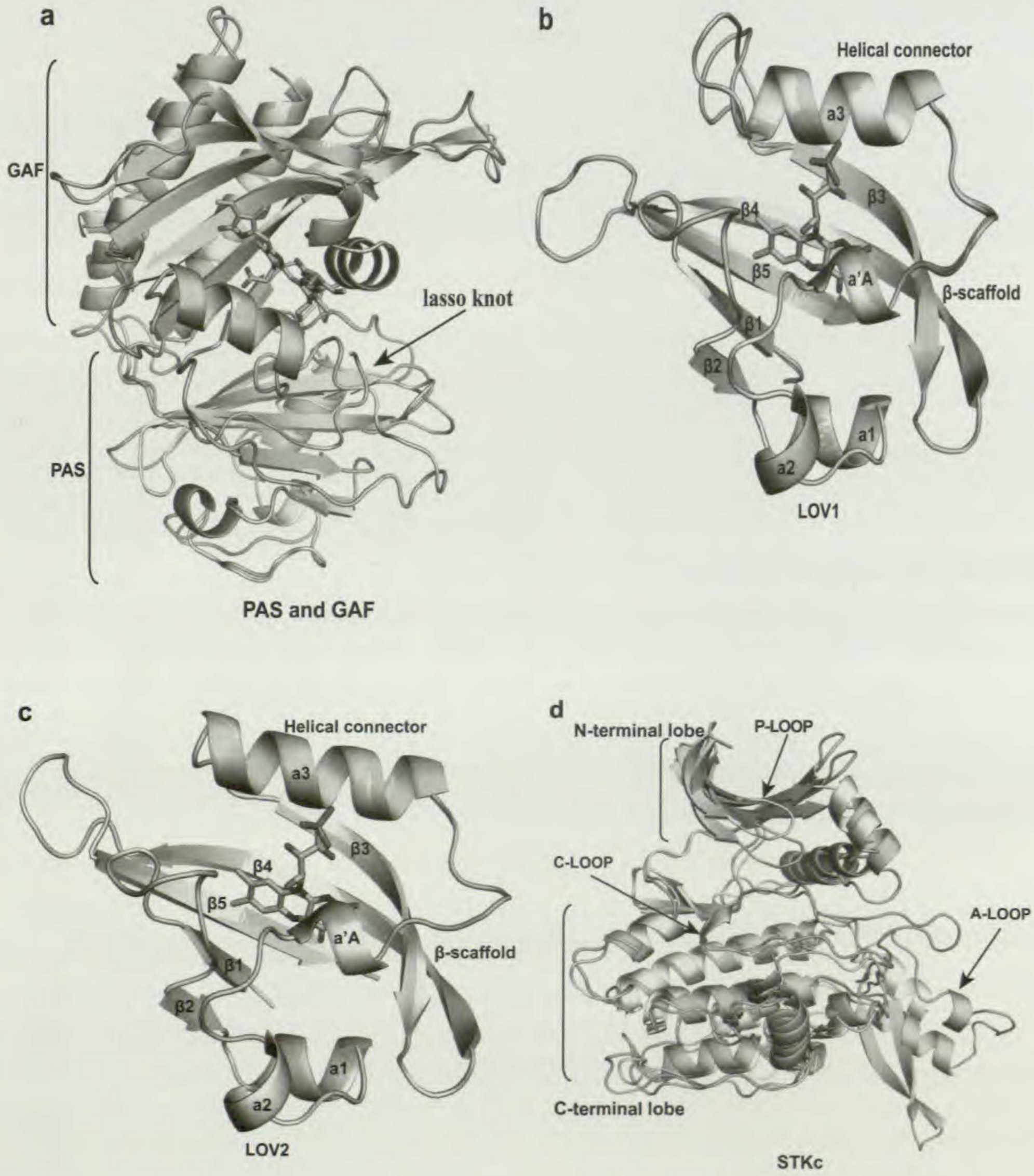

FIG. 4. The alignment of the predicted three-dimensional structure of NEO functional domains for three fern species *Coniogramme intermedia* var. *glabra* (dark gray), *Plagiogyria distinctissima* (gray) and *Pronephrium lakhimpurnense* (light gray). **(a)** The predicted 3D structure of PAS and GAF domains. The phytochromobilin (PΦB) linked with cys residues are displayed with sticks. The figure-eight knots are marked with black arrow. **(b)** and **(c)** The predicted 3D structure of LOV1 and LOV2 domains, respectively. The FMN ligand linked with cys residues are displayed with sticks. **(d)** The predicted 3D structure of STKc domains. Areas of the P-loop (phosphate-binding loop), C-loop (catalytic loop) and A-loop (activation loop) are marked with three black arrows.

including the photochemically active cysteine residues (Cys712, 982 for *Ci*NEO; Cys702, 973 for *Pd*NEO; and Cys707, 977 for *Pl*NEO). Furthermore, 11 residues that interact with FMN were identified based on the results of the flavin-binding domain (Crosson and Moffat, 2001). Residues in a′A and a3

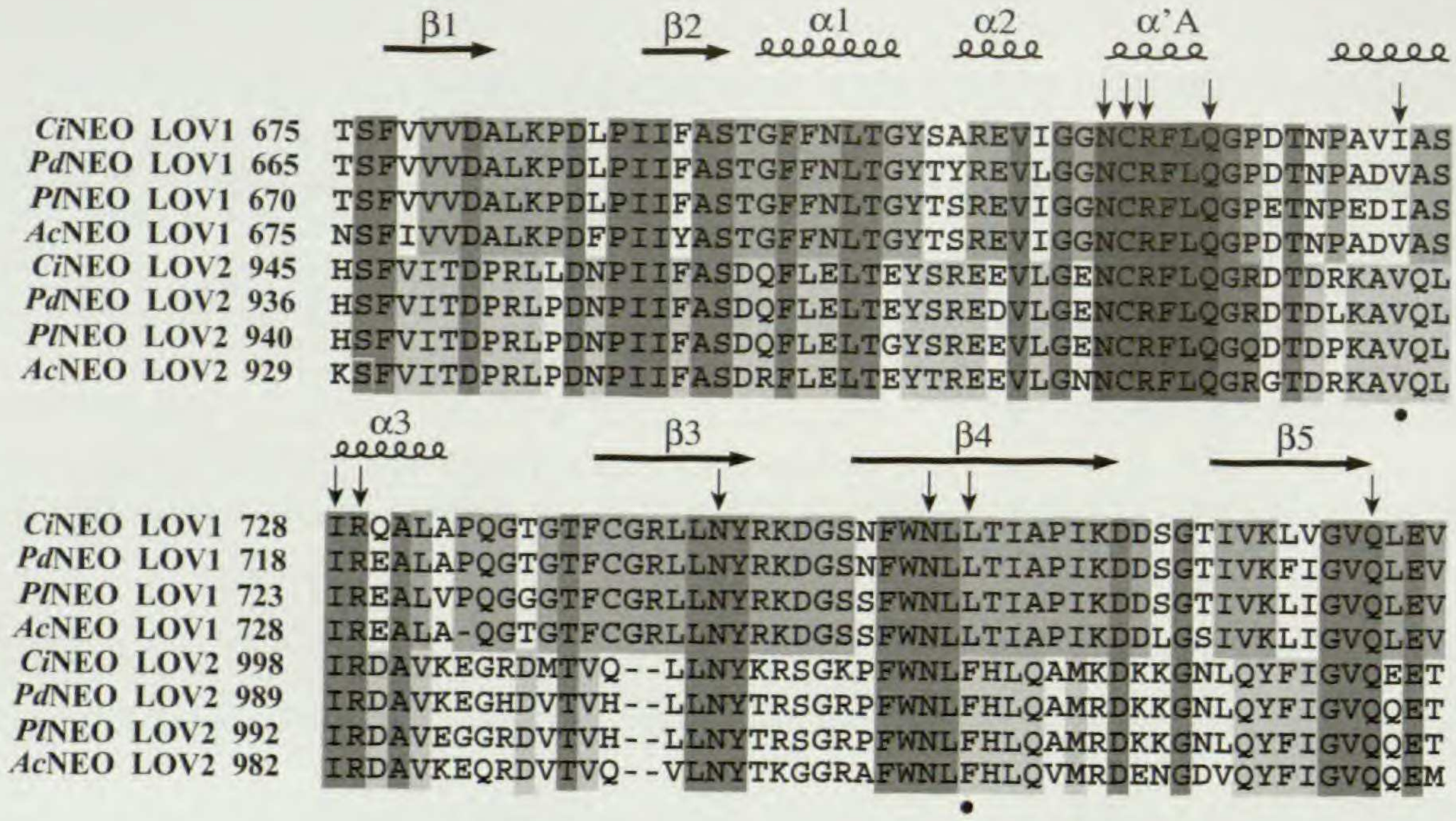

FIG. 5. Alignment of the eight LOV domains and the identification of flavin-interacting residues of *Coniogramme intermedia* var. *glabra*, *Plagiogyria distinctissima* and *Pronephrium lakhimpurnense*. NEO proteins. Secondary structure of LOV is marked above alignment: β-strand, arrows; helix, windings. Dark gray background residues are conserved in all LOV domains. Gray residues are conserved in all LOV1 domains, whereas light gray residues are conserved in all LOV2 domains. NCRFLQ residues are specific motif in LOV domains. Residues that interact with FMN are marked above with black vertical arrows, whereas variable residues interacting with FMN marked bottom with black dots.

helixes as well as β-strands 3 and 4 make the majority of flavin contacts. Site-directed mutagenesis analysis suggested that the FMN chromophore interacting site Gln1029 in *Adiantum* (corresponding to Gln1045 in the LOV2 of *Ci*NEO, Gln1036 in *Pd*NEO and Gln1039 in *Pl*NEO) plays an important role in the protein structural changes of NEO-LOV2. Most of the FMN interacting residues are conserved in the aligned LOV domains with only two exceptions. One is the residue 725 (*Ci*NEO LOV1 as reference), which is an isoleucine in *Ci*NEO and *Pl*NEO LOV1 but a valine in other LOV1 and all LOV2; the other is the residue 758, which is a leucine in LOV1 but a phenylalanine in LOV2 (Fig. 5). For *Adiantum* NEO, it has been suggested that the two amino acid differences between LOV1 and LOV2 is relative to the distinct roles that they play in the photoregulation process (Yamamoto *et al.*, 2008).

The NEO kinases of three fern species all have a α/β structure similar to PKA (c-AMP-dependent protein kinase A) containing the autophosphorylation site (Akamine *et al.*, 2003) as shown in Fig. 4d. Similarly, the NEO STKc domains all fold into a two-lobed structure including one smaller N-terminal and one larger C-terminal lobes. Based on the information of functional sites in PKA (Shaltiel *et al.*, 1998), we examined the key sites in fern NEO. The N-terminal lobe in the STKc domains contains the glycine rich and conserved motif GSGDTG as well, which contribute to the P-loop (phosphate-binding loop) of

PKA, acting as an ATP-binding site (Akamine *et al.*, 2003). The C-terminal lobe contains the typical amino acid sequences of the C-(catalytic) and the A-(activation) loop of PKA. In the C-loop regions, all the residues are conserved including the phosphorylation reaction catalyzing residue (Asp1253, Asp1244, and Asp1247 for *Ci*NEO, *Pd*NEO, and *Pl*NEO, respectively). For the A-loop, *Ci*NEO is formed mainly by two β-sheets, *Pd*NEO by two α-helixes while *Pl*NEO by one α-helixe and two small β-sheets. Whether this difference contributes to the activation ability remains unknown. By contrast, the Asp residue responsible for chelating Mg^{2+} ions and the consensus phosphorylation site (Ser1313, Ser1305 and Ser1308 for *Ci*NEO, *Pd*NEO, and *Pl*NEO, respectively), which have been reported to be required for full kinase activity (Tokutomi *et al.*, 2008), are conserved in all the NEO STKc domains. In summary, the NEO STKc domains for three fern species well preserve the PKA functional motifs, suggesting that NEO may have a molecular mechanism similar to PKA.

Conclusion

In this study, we have described the isolation and sequence analyses of full-length *NEO* genes from one tree fern and two polypod ferns. This may suggest that *NEO* genes are widely exist in many fern species (at least in tree ferns and polypod ferns). Molecular characterization of the three *NEO* genes reveals that their ORF size, GC content, basic protein properties and the predicted secondary and 3D structures are very similar to those of *Adiantum NEO*. In fact, the major structural motifs are all identical in fern NEOs, implying that the NEO may share a common mechanism for light-sensing in different fern species. The minor differences among them may relate to the individual diversification. Its N-terminus has three domains with homology to phytochrome and C-terminus contains three domains similar to phototropin. The comparisons of the PAS and GAF structures with plant phytochromes and STKc domain with PKA also reveal that the NEO well preserves the functional motifs. The analysis of the functional domains and key sites relating to both regulation and catalyzation may provide useful information on the molecular mechanism underlying the photoregulation mediated by NEO proteins.

Acknowledgments

This project was supported by the "100 Talent Project" of Chinese Academy of Sciences (Grant No.: 0729281F02), the National Natural Science Foundation of China (Grant No.: 30771763, 30170101), and the "Outstanding Young Scientist Project" of the Natural Science Foundation of Hubei Province, China (Grant No.: 0631061H01).

Literature Cited

Akamine, P., Madhusudan, J. Wu, N. H. Xuong, L. F. Ten Eyck and S. S. Taylor. 2003. Dynamic features of cAMP-dependent protein kinase revealed by apoenzyme crystal structure. J. Mol. Biol. 327:159–171.

ANTAL, Z., C. RASCLE, M. FEVRE and C. BRUEL. 2004. Single oligonucleotide nested PCR: a rapid method for the isolation of genes and their flanking regions from expressed sequence tags. Curr. Genet. 46:240–246.

BOYLAN, M. T. and P. H. QUAIL. 1991. Phytochrome a overexpression inhibits hypocotyl elongation in transgenic *Arabidopsis*. Proc. Natl. Acad. Sci. USA 88:10806–10810.

BRIGGS, W. R. and J. M. CHRISTIE. 2002. Phototropins 1 and 2: versatile plant blue-light receptors. Trends Plant Sci. 7:204–210.

CAMPBELL, T. J. and E. LISCUM. 2001. Plant photobiology 2001: a thousand points of enlightenment from receptor structures to ecological adaptation. Plant Cell 13:1704–1710.

CHRISTIE, J. M. 2007. Phototropin blue-light receptors. Annu. Rev. Plant Biol. 58:21–45.

CROSSON, S. and K. MOFFAT. 2001. Structure of a flavin-binding plant photoreceptor domain: insights into light-mediated signal transduction. Proc. Natl. Acad. Sci. USA 98:2995–3000.

DELANO, W. L. 2002. The PyMOL molecular graphics system. DeLano Scientific, Palo Alto, CA. http://www.pymol.org.

EMANUELSSON, O., S. BRUNAK, G. VON HEIJNE and H. NIELSEN. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. 2:953–971.

FOLTA, K. M. and E. P. SPALDING. 2001. Unexpected roles for cryptochrome 2 and phototropin revealed by high-resolution analysis of blue light-mediated hypocotyl growth inhibition. Plant J. 26:471–478.

GASTEIGER, E., C. HOOGLAND, A. GATTIKER, S. DUVAUD, M. R. WILKINS, R. D. APPEL and A. BAIROCH. 2005. Protein identification and analysis tools on the ExPASy server. The Proteomics Protocols Handbook. Humana Press, Totowa. 571–607.

GEOURJON, C. and G. DELEAGE. 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput. Appl. Biosci. 11:681–684.

GUINDON, S. and O. GASCUEL. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704.

HOFMANN, K., P. BUCHER, L. FALQUET and A. BAIROCH. 1999. The PROSITE database, its status in 1999. Nucl. Acids Res. 27:215–219.

HOFMANN, K. and W. STOFFEL. 1993. TMbase - A database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler 374:166.

HUANG, S. H. 1997. Inverse PCR approach to cloning cDNA ends. Methods Mol. Biol. 69:89–96.

KAGAWA, T. and M. WADA. 1994. Brief irradiation with red or blue light induces orientational movement of chloroplasts in dark-adapted prothallial cells of the fern *Adiantum*. J. Plant Res. 107:389–398.

KANEGAE, T., E. HAYASHIDA, C. KURAMOTO and M. WADA. 2006. A single chromoprotein with triple chromophores acts as both a phytochrome and a phototropin. Proc. Natl. Acad. Sci. USA 103:17997–18001.

KAWAI, H., T. KANEGAE, S. CHRISTENSEN, T. KIYOSUE, Y. SATO, T. IMAIZUMI, A. KADOTA and M. WADA. 2003. Responses of ferns to red light are mediated by an unconventional photoreceptor. Nature 421:287–290.

NOZUE, K., T. KANEGAE, T. IMAIZUMI, S. FUKUDA, H. OKAMOTO, K. C. YEH, J. C. LAGARIAS and M. WADA. 1998. A phytochrome from the fern *Adiantum* with features of the putative photoreceptor NPH1. Proc. Natl. Acad. Sci. USA 95:15826–15830.

PANDINI, A. and L. BONATI. 2005. Conservation and specialization in PAS domain dynamics. Protein Eng. Des. Sel. 18:127–137.

PEPPER, A. E., M.-S. SEONG-KIM, S. M. HEBST, K. N. IVEY, S. J. KWAK and D. E. BROYLES. 2001. shl, a new set of *Arabidopsis* mutants with exaggerated developmental responses to available red, far-red, and blue light. Plant Physiol. 127:295–304.

POSADA, D. 2003. Using MODELTEST and PAUP to select a model of nucleotide substitution. Curr. Protoc. Bioinf. Chapter 6:Unit 65.

QUAIL, P. H. 2002. Phytochrome photosensory signalling networks. Nat. Rev. Mol. Cell Biol. 3:85–93.

ROCKWELL, N. C., Y. S. SU and J. C. LAGARIAS. 2006. Phytochrome structure and signaling mechanisms. Annu. Rev. Plant Biol. 57:837–858.

SAKAI, T., T. KAGAWA, M. KASAHARA, T. E. SWARTZ, J. M. CHRISTIE, W. R. BRIGGS, M. WADA and K. OKADA. 2001. *Arabidopsis* nph1 and npl1: blue light receptors that mediate both phototropism and chloroplast relocation. Proc. Natl. Acad. Sci. USA 98:6969–6974.

SAMBROOK, J., E. F. FRITSH and T. MANIATIS. 2001. Molecular cloning: A laboratory manual. Cold Spring Harbor. Cold Spring Harbor Laboratory Press.

SCHNEIDER, H., E. SCHUETTPELZ, K. M. PRYER, R. CRANFILL, S. MAGALLON and R. LUPIA. 2004. Ferns diversified in the shadow of angiosperms. Nature 428:553–557.

SHALTIEL, S., S. COX and S. S. TAYLOR. 1998. Conserved water molecules contribute to the extensive network of interactions at the active site of protein kinase A. Proc. Natl. Acad. Sci. USA 95:484–491.

SUETSUGU, N., F. MITTMANN, G. WAGNER, J. HUGHES and M. WADA. 2005. A chimeric photoreceptor gene, NEOCHROME, has arisen twice during plant evolution. Proc. Natl. Acad. Sci. USA 102:13705–13709.

TOKUTOMI, S., D. MATSUOKA and K. ZIKIHARA. 2008. Molecular structure and regulation of phototropin kinase by blue light. Biochim. Biophys. Acta. 1784:133–142.

WADA, M. and A. KADOTA. 1989. Photomorphogenesis in lower green plants. Annu. Rev. Physio. Plant Mol. Biol. 40:169–191.

WAGNER, J. R., J. S. BRUNZELLE, K. T. FOREST and R. D. VIERSTRA. 2005. A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. Nature 438:325–331.

YAMAMOTO, A., T. IWATA, S. TOKUTOMI and H. KANDORI. 2008. Role of Phe1010 in light-induced structural changes of the neo1-LOV2 domain of *Adiantum*. Biochemistry 47:922–928.